图书情报工作
第 64 卷 第 14 期 2020 年 7 月

科学文献被引轨迹的跨学科特征影响研究*

■ 石燕青[1] 赵一方[2]

[1] 南京农业大学信息科技学院 南京 210095 [2] 南京大学信息管理学院 南京 210023

摘 要:[目的/意义]文献引用不仅是静态的统计分布,也是一个复杂的动态发展过程。对于引文曲线的变化及其影响因素进行进一步探索,有助于了解某一学科的文献生命周期和动态发展过程,同时对科学预测和科研评估也具有潜在价值。[方法/过程]对不同类型的引文轨迹进行识别和分类,研究各类型引文轨迹的特征、时间分布以及其他可能包含在引文曲线中的信息。对论文的跨学科程度进行测度,考察跨学科程度与引文轨迹之间的关系。[结果/结论]在不同的文献累积引用水平下,都存在"持续性"和"暂时性"两种引用模式。对于高被引文献,跨学科程度与其累积被引趋势并无明显关系,但是对于区分"持续性"和"暂时性"引用具有一定程度的影响作用。

关键词:引文分析 引文曲线 引用模式 跨学科特征
分类号:G250
DOI:10.13266/j.issn.0252-3116.2020.14.008

1 引言

引文轨迹是指一篇论文在发表之后一段时间内的被引频次随时间的变化趋势。传统的期刊影响因子、h指数等静态累计被引指标并不能客观地表达和衡量文献的被引水平[1],引文轨迹从文献被引的历史过程入手,能够从动态发展的角度衡量论文的质量、反映论文的影响力[2]。一般认为,文献在发表之后的 2-6 年达到引用顶峰,随后开始进入衰退期[3]。然而,实际中的引文轨迹发展模式在引文生命周期内表现出很大的差异。1979 年,A. Avramescu 总结出 5 种引文轨迹曲线:根据峰值高、中、低程度不同的 3 种经典引文曲线、一直单调递增的"天才型"引文曲线、前期被认可但又突然被摒弃的"昙花一现"型引文曲线[4]。李江等对诺贝尔奖得主的论文进行拟合,总结出 5 种与前人研究有差异的引文轨迹:经典型、指数增长型、睡美人型、双峰型、波型[5]。将引文时间跨度拉长,学者们还发现了一类在发表之初被忽视、多年后被引量突增的"睡美人"文献[6],与之相反的是"昙花一现"型文献[7-9]。

文献引用轨迹的变化与引文网络的动态变化密切相关,对于单篇论文而言,其引文轨迹的发展可能会受到学科差异[10]、作者特征[11-12]、文献类型[13-14]、论文篇幅[15]等多种因素影响[16-17],E. Roldan-Valadez 将这些因素归纳为 3 类:论文相关因素、期刊相关因素、作者相关因素[18]。在诸多因素中,科研成果的跨学科特征不仅可以反映出知识在不同学科间的流动情况,同时,科研成果的跨学科程度往往与其影响力有密切关系[19]。在已有研究中,不同学者对跨学科特征和文献被引的关系得出了不同的结论。J. M. Levitt[20]、S. Kwon[21]等以社会科学文献为研究对象,G. Abramo 以自然科学文献为对象,均证实了跨学科特征与被引的正相关关系[22];H. S. Bhat 同时研究了自然科学和社会科学文献,同样发现了跨学科性对被引的正向影响作用[23]。而 I. V. Pomomarev[24]、E. J. Rinia[25]分别以化学、物理作为研究领域,发现跨学科特征与被引间的关系不显著。此外,在 J. M. Levitt 和 M. Thelwall[26]、J. Wang[27]等的研究中,跨学科特征与被引表现出一定的负相关关系。

可以看出,跨学科特征对文献被引的影响作用非常复杂,由于测度指标、研究对象的多样化,研究结论也呈现出差异性。此外,现有研究大多关注跨学科性对文献累积被引水平的影响,鲜少对跨学科特征与引文轨迹的关系进行探讨。基于此,本文以"Library and Information Science"学科的文献为研究对象,从引文轨

* 本文系国家自然科学基金项目"引文扩散理论及实证研究"(项目编号:71874077)研究成果之一。
作者简介:石燕青(ORCID:0000-0002-1091-7926),讲师,博士,E-mail:yqs4869@njau.edu.cn;赵一方(ORCID:0000-0001-8675-9322),硕士研究生。
收稿日期:2019-11-25 修回日期:2020-03-27 本文起止页码:74-84 本文责任编辑:王传清

迹分布的角度来考察论文的被引轨迹模式及其特征，并考察文献的跨学科程度与引文轨迹之间的关系，一方面有助于我们更好地理解知识传播的过程，另一方面为文献被引情况预测提供参考。尽管本文从图书情报学科视角进行文献的引文分析，文中所用的分析框架和方法可以在其他层面和视角（如期刊、机构、国家、研究主题等）进行应用和延伸。

2 研究设计与数据收集

图书情报学科具有一定的跨学科性[28]，与经济学、管理学、计算机科学等多个学科都具有一定程度的亲缘关系，因此，该学科文献的被引程度较高，学科引用关系足够多样化，便于进行论文的跨学科特性研究。本文以图书情报学科为研究领域，首先，从时序角度对文献被引轨迹的变化过程进行刻画，通过模型拟合对不同类型的引文轨迹进行分类；在此基础上，以单篇文献的跨学科程度为自变量，以文献被引类型为因变量，考察文献的跨学科特征与其被引情况的关系。

2.1 引文轨迹研究方法概述

对于引文轨迹的刻画，指数分布、逻辑分布、Weibull 分布、Gamma 分布、Beta 分析等多种模型都被应用于引文曲线的建模研究中[29-33]。为了考察时间维度下文献被引频次的分布情况以及不同的引文模式，本文采用基于亚组的轨迹分析模型（Group-Based Trajectory Modeling, GBTM）[34]对引文轨迹进行分类，并研究相应的引文曲线特征。GBTM 是心理学领域中研究个体行为随时间动态发展的一种常用方法，基于数据间固有的异质性，识别能够表征数据变化趋势的轨迹组数和形状，从而将个体分配到最合适的组中。每一组曲线并非数据变化的精确轨迹，而是能够相对准确地描述数据集中个体发展异质性的可视化表示[35]。

建立分组模型的过程如下：

首先，通过逐步增加分组数量来确定组数，并通过贝叶斯信息准则（Bayesian Information Criterion, BIC）来选择能够最大程度地表达组间异质性的模型，BIC 准则通常被用来衡量模型的拟合优良性并兼顾模型的简洁度。针对模型复杂度增加而导致的过拟合现象，以及样本数量过多时模型精度过高造成的模型复杂度过高问题，BIC 引入了与模型参数个数相关、同时考虑样本数量的惩罚项：$BIC = -2\ln(L) + k\ln(n)$，其中 L 是对数似然值，k 是参数个数，n 是样本数量，BIC 值越接近于 0 表示模型拟合越好。

确定了引用轨迹组数后，根据经验并结合数据的实际情况来判断确定曲线拟合的形状。从文献的引用模式来看，大多数期刊文献在发表之后第 2-3 年达到引用峰值[36]，之后引用次数逐渐下降并保持平稳。因此，在本研究中，首先将所有组的初始曲线设为三次多项式方程，在分析过程中再根据曲线的结果和其他指标来逐步调整。

确定组数和曲线方程之后，根据模型中每个组的平均后验概率（average posterior probabilities, APP）来验证分组的准确性。对于有 K 组的模型，每篇文章都有 L 个对应的后验概率，根据最大概率的分组原则，每篇文章会被分到相应的组中，对该分组中所有文章的后验概率取平均值，即为该组的平均后验概率。D. S. Nagin[34-35]指出，在分组结果中，每个组的平均后验概率超过 70%，表明该组内所有数据都得到了较好的分组。需要指出的是，在拟合过程中，上述各个步骤可以交替进行，根据实际的结果返回上一步骤进行调整。

2.2 论文跨学科程度及其测度

为了研究论文的跨学科程度与引文轨迹的关系，本文对论文的跨学科程度进行测量，并通过识别不同的引文轨迹曲线，考察论文的跨学科程度是否会对论文的引用模式产生影响。

以单篇论文为对象，本文从两个角度测度了论文的跨学科程度。借鉴专利引证分析中的“前向”引用关系和“后向”引用关系，以及 A. L. Porter 提出的学科集成指数和学科扩散指数模型[37-38]，将论文的“前向”引用关系，即施引文献看作知识的扩散过程，论文在各个学科的扩散程度用“学科扩散指数”来测度；将论文的“后向”引用关系，即参考文献看作知识的集成过程，论文对各个学科的集成化程度用“学科集成化指数”来测度。跨学科程度，即学科的多样化程度测量借鉴了 A. Stirling 提出的多样化测度模型[39]。在该模型中，多样性程度的量化分析应该综合考察 3 个基本特性：多样性（variety），即系统中元素所属的类目数；均衡性（balance），即系统中的元素在每个类中分布的均衡程度；异质性（disparity），即每个类之间的差异性。结合这 3 个基本元素的多样化计算公式为：

$$D = \sum d_{ij} \cdot p_i \cdot p_j \qquad \text{公式(1)}$$

其中，p_i 和 p_j 分别是 i 和 j 在系统中所占的比例（均衡性），d_{ij}是 i 和 j 之间的差异程度（异质性），求和体现出系统的多样性。

基于该模型，本文对两个跨学科程度的指标进行测度：① 学科集成化指数（Integration, INT）。将单篇

论文的参考文献集作为一个系统，计算参考文献集的跨学科程度。以不同学科间的共现次数作为学科间的距离 d_{ij}，即学科间的异质性特征，以某一学科在参考文献集的所有学科中所占的比例来衡量系统的均衡性特征，依据公式(1)计算参考文献集的学科多样化程度。② 学科扩散化指数(Diffusivity, DIF)。将单篇论文的施引文献集作为一个系统，计算施引文献集的跨学科程度。学科扩散化指数的计算方法与学科集成化指数相同。

2.3 数据来源

本文数据来源于 Web of Science 数据库，以 Web of Science 学科分类为依据，选择 Library and Information Science 作为研究领域。以 10 年作为引文轨迹发展的时间窗口，考察论文自发表之后的引文模式和变化趋势。在 Web of Science 中检索“Library and Information Science”学科分类下 2003 – 2007 年的文献，为了排除不同的文献类型(如综述、报告、文献)间引用模式的差异性，本文限定文献类型为“article”，检索结果共得到 10 406 条文献数据，剔除 1 884 条零被引文献之后，剩余 8 522 条数据记录。

这些文献的被引频次离散程度过大且分布不均，60% 以上年均被引频次小于 1。以单篇论文为对象时，拟合过程中部分引文轨迹的趋势特征会被论文间过大的被引趋势差距所削弱，掩盖大量有价值的信息，使得统计分析的准确性大大降低，也为曲线拟合带来困难。为了更细粒度地分析引文轨迹曲线，减少数据离散程度对分析效果的影响，依据帕累托法则将文献进行了层次划分并分别对其进行分析：低被引文献(年均被引频次 ≤1)5 258 篇、中被引文献(1 < 年均被引频次 ≤12)3 123 篇、高被引文献(年均被引频次 > 12)141 篇。其中，低被引文献的年均被引频次不足 1，曲线趋势变化程度过小，因此不作为本文的研究对象，分别对 3 123 篇中被引文献和 141 篇高被引文献进行分析。

3 基于亚组模型的文献被引轨迹分析

3.1 高被引文献的引文轨迹分析

首先，将分组范围确定为 3 – 8 个组，使用三次多项式方程对 141 篇高被引文献的引文曲线进行拟合，发现 4/5/7/8 组均拟合失败。尝试将回归方程的幂次增加为四次方，发现 8 组仍无法拟合。将幂次增加为五次方时，所有组别均可以成功拟合，3 – 8 组的拟合方程 BIC 值分别为 – 5 555.32、– 5 488.55、– 5 421.36、– 5 379.38、– 5 326.73、– 5 300.14，说明模型的拟合优度有所提升，曲线拟合结果见图 1。

从拟合结果来看，随着组数的增加，拟合趋势曲线不断被细分，出现越来越多的曲线类型。同时 BIC 值增大，表明方程拟合效果越来越好。然而，对比拟合图形发现，当预设组数超过最佳组数时，新增类目的表征信息较少，弱化了分组分类的差异性，当组数增加到 7 组时，曲线的细分只是发生在引用次数较低的文献中，即从 6 组到 7 组的变化中，图 1(d)中的 3 条曲线(第二组(24.1%)、第三组(63.4%)、第四组(7.3%))分化为图 1(e)中的 4 条曲线(第二组(34%)、第三组(49.2%)、第四组(8.7%)、第五组(2.9%))，从整体趋势来看，并未出现新的曲线形状，因此选择包含所有代表性曲线且更为简洁的 6 组拟合结果(即图 1(d))进行分析。

由图 1(d)可以看出，第二组(24.1%)和第三组(63.4%)中的文献占比超过 80%，且这两组的引文曲线形状很接近，在发表后很长一段时间保持较低且稳定的被引水平，尤其在第三组中，文献的年均被引次数不到 20 次。

相比之下，第四组中的文献被引次数则保持了相对稳定的增长趋势，每年的平均被引次数保持在中等偏下水平，且随着时间发展，曲线增长趋势后继有力，在所有高被引文献中占比不到 10%，属于影响力中等的文献。

其余 3 条曲线中，第六组的总体被引水平最高，在发表 10 年内始终保持稳定的增长趋势，保持持续稳定的高影响力，属于同时具有高质量和高影响力的一类文献。第一组和第五组的文献累计引用次数相近，但曲线趋势却完全不同：在第一组中，文献在发表之后的前 5 年被引次数偏低，但仍保持着低速增长的趋势，在第五年之后引用次数快速上升，在第七年之后年被引次数超过 100 次；而在第五组中，文献在发表之初被引次数快速增长，并在发表后的第 3 – 4 年达到顶峰，随后被引趋势逐渐衰减，出现该被引趋势的文献仅占比 0.7%(1 篇)，在年均被引次数超过 12 次的高被引文献群体中，先增后降的文献类型占比较小。1985 年，E. S. Aversa 通过文献聚类发现了“延迟增长 – 缓慢下降”和“立即增长 – 快速下降”两种引文趋势[40]。S. E. Baumgartner 和 L. Leydesdorff 对 6 种期刊的文献进行了引文轨迹分析，同样发现了 E. S. Aversa 提到的两类曲线[41]。相比于第五组的“暂时性”引用，第一组“持续性”引用曲线的波峰一般出现得较晚，且两者波

石燕青，赵一方. 科学文献被引轨迹的跨学科特征影响研究[J]. 图书情报工作,2020,64(14):74－84.

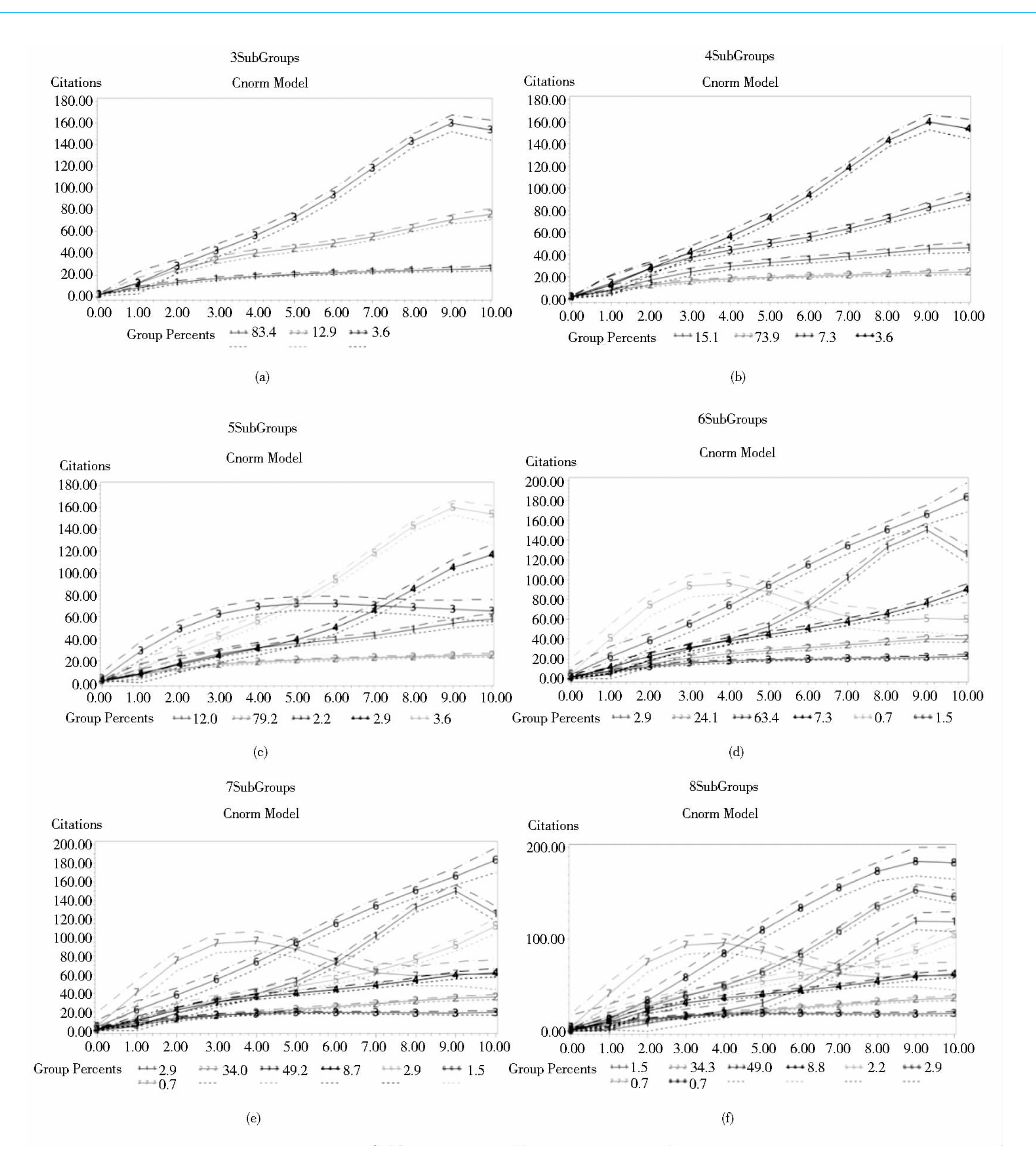


图 1　高被引文献引文轨迹 3－8 组分组结果

峰出现的时间差因学科、文献类型等有差距，因此在较短的时间窗口内难以准确地衡量文献的质量或影响力。

3.2　中被引文献的引文轨迹分析

与高被引文献的分析方法类似，使用 5 次幂方程对 3 124 篇高被引文献进行拟合，对比不同组数的拟合结果发现，随着细分的组数增多，方程的 BIC 值逐渐增加。同样，在分析引文趋势时我们更加关心被引频次较高的文献，因此在模型的拟合过程中逐渐增加组数，直到不再出现相对较高被引文献细分的组别。图 2 展示了 3－6 个组的拟合结果，相比 4 组模型（图 2（b）），5 组模型中出现了新的代表性曲线，而相比 5 组模型（图 2（c）），6 组模型（图 2（d））只是在低被引文献中进行了细分，且并没有新趋势出现。因此，选择 5 组模型（2（c））作为分析对象。

从图 2（b）中可以看出，在第一组中，60% 以上的文献在发表之后 10 年内年均被引频次不超过 2 次，第二组的引文趋势与第一组类似，这两组文献占了总文献的 80% 以上，这一情况与高被引文献的结果类似。同样，在中被引文献中，也发现了"持续性"引用（第四组和第五组）和"暂时性"引用（第三组）。第三组中的文献在发表后的第 2－3 年达到被引峰值，这类文献往往是关注相关领域热点的研究，在发表之初引起较为广泛的关注，但随后随着新的理论或主题的出现，被引频次逐渐下降，而第四组的引文曲线保持稳定的上升趋势。第五组的文献从发表之后就保持持续上升的被引频次，是所有文献中影响力最高的一类。

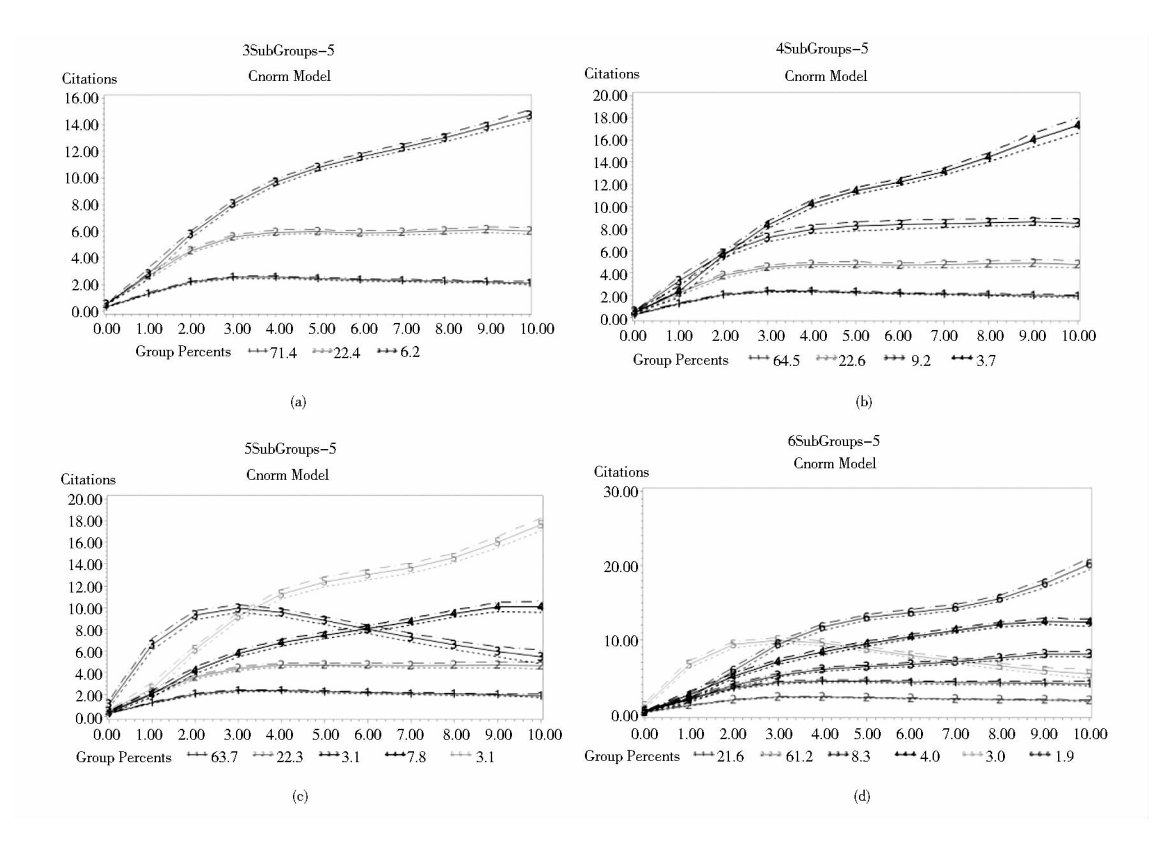


图 2　中被引文献引文轨迹 3 - 6 组分组结果

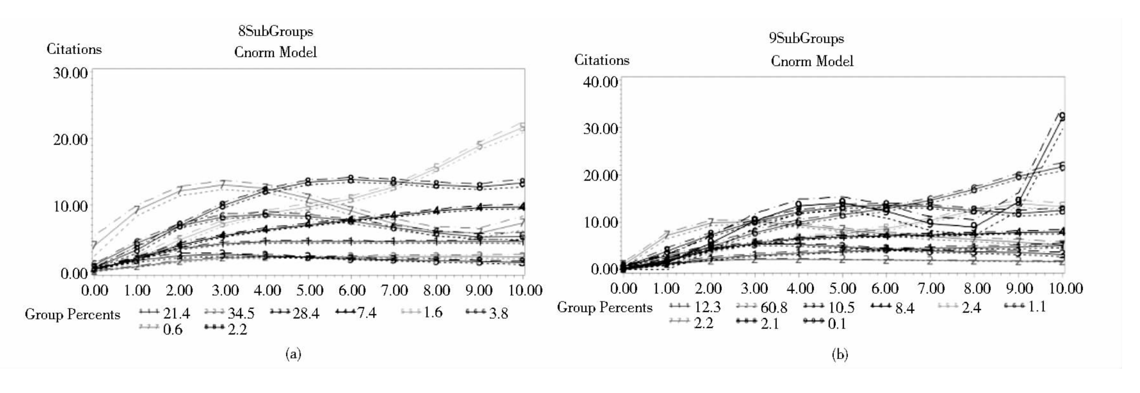


图 3　中被引文献引文轨迹 8 - 9 组分组结果

随着组数增加到 8 组，曲线中出现了更多“持续性”引用和“暂时性”引用的组合。图 3(a) 是组数为 8 组的分组结果，从图中可以看到，累积被引次数较高的第七组(0.6%)和第八组(2.2%)、累积被引次数较低的第六组(3.8%)和第四组(7.4%)都呈现出“暂时性”和“持续性”引用的特点。与图 2(b) 中的“暂时性”被引文献一样，第六组和第七组文献也在发表之后第三年达到被引峰值，随后被引频次开始下降。当组数增加为 9 组(图 3(b))时，出现了一条新的被引轨迹，即图 3(b) 中的第九组文献(0.1%)，在前 8 年表现出经典引文曲线的特征，但在生命周期即将结束时，被引趋势快速上升，开始了第二次生命周期，随着第二次进入衰退期，其趋势将发展为文献[6]中的“双峰曲线”。

4　跨学科特征对文献被引轨迹的影响作用分析

从引文轨迹分析的结果中，不仅能够区分出论文的整体引用水平，还能够区分出论文的影响力持久度。基于引文轨迹的分组结果，本文使用美国经济学家 D. McFadden 提出的 Multinomial Logit (MNL) 模型[42-43]探究文献特征与引文模式间的关系。MNL 模型使用最大似然估计法来估计“类别特质”或“个体特质”对类别选择的影响，模型理论坚

石燕青，赵一方. 科学文献被引轨迹的跨学科特征影响研究[J]. 图书情报工作,2020,64(14):74－84.

实而计算简单,是分析多类型因变量的影响因素的经典模型。本文使用该模型来研究论文出现在各组文献集中的概率,以及论文被引模式与其跨学科程度的关系。

如前所述,以单篇论文的分组结果作为因变量,论文的跨学科程度作为自变量,使用两个指标:论文的学科集成化指数(Integration, INT)和学科扩散化指数(Diffusivity, DIF)来表示。同时,将论文的作者数量(Number of Authority, AUN)、参考文献数量(Number of references, REN)、论文篇幅(Pages, PG)、文献所在期刊的影响因子(Impact Factor, IF)、国际合作程度(CON)、标题长度(TIL)作为控制变量。

4.1 高被引文献的跨学科影响

在高被引文献的6组分类(图1(d))结果中,以第五组引文轨迹作为参考组进行模型拟合,结果如表1所示。每一组的结果解释为:与参考组相比,引文轨迹落入该组的可能性。系数Coef.表示自变量对因变量作用大小的度量;比数比OR是流行病研究中的概念,用来衡量疾病与暴露之间的联系强度,OR > 1时自变量具有正向作用,OR < 1表示自变量具有反向作用。在本研究中,该指标的含义可以解释为:该引文轨迹组与参考组的相对危险度是OR,即相比于参考组,自变量每增加一个单位,文献落入该引文轨迹组的概率增加(OR－1)倍或减少(1－OR)%。

表1 高被引文献的6组引文轨迹多项式逻辑回归结果

		Coef.	Std. Err.	P > |z|	OR	95% Conf. interval for OR	
Group 1	Intercept	－118.23 ***	2.41	0.000	4.49E－52	3.94E－54	5.12E－50
	AUN	42.01 ***	0.29	0.000	1.75E＋18	9.98E＋17	3.10E＋18
	IF	－4.95 ***	0.09	0.000	0.01	5.94E－03	8.53E－03
	REN	－0.79	0.92	0.197	0.46	0.075	2.77
	PG	2.74 ***	0.44	0.000	15.44	6.57	36.3
	CON	158.43 ***	0.99	0.000	6.42E＋68	9.19E＋67	4.48E＋69
	TIL	3.13 ***	0.75	0.000	22.88	5.22	100
	DIF	133.92 ***	0.54	0.000	1.44E＋58	5.05E＋57	4.12E＋58
	INT	－311.88 ***	2.51	0.000	3.55E－138	2.61E－138	4.83E－134
Group 2	Intercept	－116.99 ***	2.22	0.000	1.56E－51	2.01E－53	1.21E－49
	AUN	42.07 ***	0.28	0.000	1.86E＋18	1.07E＋18	3.25E＋18
	IF	－4.92 ***	0.09	0.000	0.01	6.15E－03	8.64E－03
	REN	－0.80	0.92	0.192	0.45	0.074	2.73
	PG	2.70 ***	0.44	0.000	14.87	6.33	34.9
	CON	158.64 ***	0.97	0.000	7.85E＋68	1.18E＋68	5.22E＋69
	TIL	3.13 ***	0.75	0.000	22.94	5.23	101
	DIF	133.71 ***	0.50	0.000	1.17E＋58	4.36E＋57	3.15E＋58
	INT	－311.24 ***	2.36	0.000	6.74E－136	6.56E－138	6.92E－134
Group 3	Intercept	－121.00 ***	5.28	0.000	2.81E－53	8.94E－58	8.88E－49
	AUN	41.65 ***	0.44	0.000	1.23E＋18	5.21E＋17	2.90E＋18
	IF	－4.91 ***	0.16	0.000	0.01	5.34E－03	1.01E－02
	REN	－0.77	0.92	0.203	0.46	0.076	2.83
	PG	2.88 ***	0.44	0.000	17.82	7.5	42.3
	CON	159.93 ***	1.25	0.000	2.86E＋69	2.46E＋68	3.33E＋70
	TIL	3.11 ***	0.75	0.000	22.43	5.11	98.3
	DIF	135.30 ***	1.05	0.000	5.76E＋58	7.40E＋57	4.49E＋59
	INT	－317.21 ***	5.40	0.000	1.73E－138	4.37E－143	6.83E－134
Group 4	Intercept	－116.97 ***	3.42	0.000	1.59E－51	1.96E－54	1.29E－48
	AUN	40.41 ***	0.55	0.000	3.53E＋17	1.20E＋17	1.04E＋18
	IF	－4.85 ***	0.11	0.000	0.01	6.33E－03	9.74E－03
	REN	－0.78	0.92	0.199	0.46	0.075	2.79
	PG	2.77 ***	0.44	0.000	16.00	6.78	37.8
	CON	160.39 ***	1.09	0.000	4.55E＋69	5.38E＋68	3.85E＋70

图书情报工作
第 64 卷 第 14 期 2020 年 7 月

（续表 1）

		Coef.	Std. Err.	P > |z|	OR	95% Conf. interval for OR	
	TIL	3.11 ***	0.75	0.000	22.38	5.11	98.1
	DIF	134.93 ***	0.69	0.000	3.99E+58	1.03E+58	1.54E+59
	INT	-314.52 ***	3.49	0.000	2.56E-137	2.76E-140	2.37E-134
Group 6	Intercept	143.55 ***	3.52	0.000	2.21E+62	2.23E+59	2.20E+65
	AUN	41.30 ***	0.90	0.000	8.62E+17	1.47E+17	5.06E+18
	IF	-4.72 ***	0.20	0.000	0.01	6.00E-03	1.32E-02
	REN	-0.77	0.92	0.203	0.47	0.076	2.84
	PG	2.74 ***	0.45	0.000	15.45	6.44	37.1
	CON	-104.11 ***	3.52	0.000	6.10E-46	6.14E-49	6.06E-43
	TIL	3.12 ***	0.75	0.000	22.74	5.18	99.7
	DIF	134.72 ***	1.42	0.000	3.22E+58	1.99E+57	5.20E+59
	INT	-315.35 ***	6.30	0.000	1.11E-137	4.78E-143	2.58E-132

注：“ *** ”表示在 0.001 水平下显著；“ ** ” 表示在 0.01 水平下显著；“ * ” 表示在 0.05 水平下显著；“ · ”表示在 0.1 水平下显著；“ ” 表示不显著

图 1(d) 中，第五组是唯一的一组“暂时性”被引文献，与之对应，除了“持续性”引用特征明显的第一组，第四组和第六组也可以看作“持续性”引用。从表 1 的结果中可以看出，DIF 和 INT 与引文轨迹有显著的关系。DIF 与文献的长期被引概率成正比，即学科扩散程度越高，论文的影响力越持久；这说明暴露在更广的学科范围内有利于论文持续受到关注，而在较窄的学科范围内，论文的影响力更容易随时间衰落。此外，此类论文可能提出了在多个领域适用的普适性方法或结论，此类研究更容易成为长久不衰的科学成果。另一方面，INT 与文献的影响力持久度成反比，即参考文献的跨学科程度越大、学科集成化指数越高，文献被“暂时性”引用的概率越大；吸收了多学科知识的研究以及本身跨学科程度高的论文，其研究深度可能因为知识跨度的增加而受限，因此在研究型论文中，此类成果的被引概率可能低于同领域中研究深度更深的成果。

进一步地，以第六组引文轨迹作为参考组进行拟合，结果显示，在第五组的各项指标中，DIF（β = -15.92，p = 0.000）和 INT（β = 254.47，p = 0.000）是显著的，DIF 与文章的影响力持久度成正比，而 INT 与影响力持久度成反比，与前一次的分析结论一致。在第一组 ~ 第四组的指标中，DIF 与 INT 指标均不显著，说明论文的总体被引水平高低并不受到学科多样化程度的影响。

综合两次拟合的结果来看，学科跨度指标对于文献的整体累计引用水平影响并不显著；论文在其他学科中的扩散程度越高，论文的影响力越持久，而论文的学科集成化程度，即参考文献中包含的学科多样化程度却对论文的影响力持久性产生负向影响。

4.2 中被引文献的跨学科影响

对于中被引文献，以 5 组分组结果为对象，以第三组作为参考组进行多项式拟合，结果如表 2 所示。在图 2(c) 的引文轨迹中，第三组属于“暂时性”引用，第四组和第五组属于“持续性”引用，从表 2 的结果来看，在区分这两类文献方面，DIF 和 INT 均没有显著影响。

进一步地，选择第五组作为参考组进行多项式拟合，并对结果进行分析。拟合结果显示，与第五组（“持续性”引用）对应的第三组（“暂时性”引用）中，DIF 与 INT 仍不显著，与上文结果一致，这说明与高被引论文不同，在中被引水平的论文中，论文的跨学科程度与论文影响力持久度并无显著关系。在区分整体被引水平方面，在第一组和第二组的结果中，DIF（β = -0.43，p = 0.010；β = -0.34；p = 0.041）表现出显著的负向影响作用，即论文的学科扩散指数越小，论文越容易落入低被引水平的组别（即第一组和第二组）。这说明论文的学科扩散范围越广，论文的整体被引水平越高，但该指标在区分整体水平接近的两组文献（第四组和第五组）时，并未表现出显著性，这说明虽然该指标在提升文献的被引水平方面存在积极的作用，但作用范围和程度有限。

表 3 列出了所有指标对高被引文献和中被引文献引文轨迹的影响作用。

从表 3 中可以看出，影响因子作为衡量静态累计被引频次的传统指标，对论文持续被引用具有促进作用，但是在论文整体被引水平方面，只在中被引的文献中体现出促进作用，对于高被引文献并未表现出显著

表 2　中被引文献的 5 组引文轨迹多项式逻辑回归结果

		Coef.	Std. Err.	P > |z|	OR	95% Conf. interval for OR	
Group 1	Intercept	0.28	0.55	0.303	1.33	0.45	3.93
	AUN	－0.07 ·	0.05	0.088	0.94	0.85	1.03
	IF	－0.07	0.06	0.136	0.93	0.82	1.06
	REN	0.01 ·	0.01	0.064	1.01	1.00	1.02
	PG	0.06 **	0.02	0.002	1.06	1.02	1.10
	CON	－0.11	0.22	0.305	0.90	0.59	1.37
	TIL	0.01 * *	0.00	0.004	1.01	1.00	1.02
	DIF	－0.03	0.18	0.438	0.97	0.68	1.39
	INT	0.12	0.26	0.323	1.13	0.67	1.90
Group 2	Intercept	3.13 ***	0.53	0.000	22.78	8.06	64.36
	AUN	－0.17 ***	0.05	0.000	0.84	0.76	0.92
	IF	－0.25 ***	0.06	0.000	0.78	0.68	0.88
	REN	0.00	0.01	0.388	1.00	0.99	1.01
	PG	0.04 *	0.02	0.028	1.04	1.00	1.08
	CON	－0.09	0.21	0.325	0.91	0.61	1.37
	TIL	0.01 **	0.00	0.006	1.01	1.00	1.02
	DIF	－0.13	0.17	0.235	0.88	0.63	1.24
	INT	－0.01	0.26	0.477	0.99	0.60	1.62
Group 4	Intercept	－0.84 ·	0.62	0.089	0.43	0.13	1.46
	AUN	－0.16 **	0.06	0.006	0.85	0.75	0.97
	IF	0.03	0.07	0.305	1.03	0.91	1.18
	REN	0.02 **	0.01	0.006	1.02	1.00	1.03
	PG	0.06 **	0.02	0.004	1.06	1.01	1.10
	CON	－0.33 ·	0.25	0.093	0.72	0.44	1.17
	TIL	0.01 *	0.00	0.019	1.01	1.00	1.02
	DIF	0.07	0.21	0.360	1.08	0.72	1.61
	INT	0.07	0.30	0.401	1.08	0.60	1.92
Group 5	Intercept	－2.94 ***	0.74	0.000	0.05	0.01	0.23
	AUN	－0.08	0.08	0.151	0.92	0.80	1.07
	IF	0.13 *	0.07	0.031	1.14	0.99	1.31
	REN	0.02 **	0.01	0.002	1.02	1.01	1.04
	PG	0.06 **	0.02	0.008	1.06	1.01	1.11
	CON	0.20	0.27	0.222	1.23	0.73	2.07
	TIL	0.01	0.01	0.143	1.01	1.00	1.02
	DIF	0.31	0.25	0.106	1.36	0.84	2.20
	INT	－0.22	0.36	0.275	0.80	0.39	1.64

注:“ ***”表示在 0.001 水平下显著;“ **” 表示在 0.01 水平下显著;“ *” 表示在 0.05 水平下显著;“ · ”表示在 0.1 水平下显著;“ ” 表示不显著

表 3　各指标对引文轨迹的影响情况

指标	文献累计被引水平		文献影响力时长	
	高被引文献	中被引文献	高被引文献	中被引文献
作者数(AUN)			+	
影响因子(IF)		+	－	
参考文献数(REN)		+		+
论文篇幅(PG)			+	+
国际合作程度(CON)	－	+	+	
标题长度(TIL)			+	
学科扩散化(DIF)		+	+	
学科集成化(INT)			－	

的影响力;参考文献的数量在一定程度上能体现一篇论文吸收知识的广度,在中被引文献中,参考文献数对论文的整体被引水平和影响力时长都具有正向影响,但对于高被引文献,参考文献数与此两者的关系并不明显;论文篇幅对于中被引文献的影响力持久性有正向促进作用。

跨学科程度对高被引文献和中被引文献的影响作用存在差异性。对于高被引文献,论文扩散的学科范围越广,其影响力越持久;而集成较多学科知识的论文,更容易在短期内得到关注。对于中被引文献,论文影响的学科范围越广,其累计被引水平越高;论文的跨

学科程度与其是否持续被引用并无明显的关系。

5 结果讨论与总结

5.1 研究结论

本文使用基于组的轨迹分析模型,根据被引趋势将文献进行了分组,揭示了文献在发表之后十年内的不同被引模式,并在此基础上,探讨了跨学科程度与引文轨迹的关系,得出以下结论:

(1)根据以往的研究,文献在发表之后的引文轨迹大多遵循先上升后下降的趋势,符合三阶多项式曲线形状。因此,在刻画引文曲线的过程中,本文首先使用三阶多项式进行拟合,但结果显示多组数据拟合失败,而五阶多项式在所有被引水平的文献组中都取得了较好的拟合结果。说明在实际中,不论对于高被引文献还是中被引文献,在较长的时间窗口中,被引曲线都可能出现多个拐点,被引模式的动态变化是一个较为复杂的过程。

(2)从文献的被引曲线中,发现了两类被引模式:"持续性"引用和"暂时性"引用。前者在到达引用顶峰后,被引趋势下降幅度较小,表现出持续的影响力,发展到 10 年之后,仍然可以获得较高的被引频次;"暂时性"被引文献在发表之初,关注度快速上升,在第 2 -4 年达到被引顶峰,随后下降趋势明显,表现出短期的高影响力,这类文献大多是研究前沿或热点文章。

从高被引文献和中被引文献的亚组分析结果可以看出,"持续性"引用和"暂时性"引用现象在各个被引层次的文献中都存在,是一个普遍现象。因此,在评价文献影响力时,应考虑较长时间内的文献被引情况。从经验来看,"暂时性"引用的出现,大多是由于其研究主题是当时的研究热点,随着主题热度的褪去,论文的被引频次也逐渐下降。更加具体的原因以及引用系统变化的动力学机制,还需要在研究中结合论文内容和相关领域的研究趋势来进一步探究。

(3)作者数量对于文献的整体被引水平和文献影响力持久度均不产生影响;影响因子对于中被引文献的整体被引水平、影响力持久度以及高被引文献的影响力持久度产生正向影响,而对高被引文献的整体被引水平并无明显的影响作用;对于中被引文献,参考文献的数量与论文的整体被引水平和被引持久度都具有显著的正相关关系,而对于高被引文献,参考文献的数量与论文的被引情况并无明显的关系;文献篇幅对于其被引情况几乎无影响作用。

(4)论文的跨学科程度对引文轨迹有一定程度的影响。跨学科程度与论文累积被引水平的关系仅体现在中被引文献中。而在区分"持续性"引用和"暂时性"引用方面,跨学科程度的作用仅体现在高被引文献中,在年均被引次数超过 12 次的论文中,论文扩散的学科范围越广,其越容易受到持续的引用,而论文集成的学科知识越多,其越容易在短期内获得关注,后续引用量有所下降。

5.2 本文的研究局限

本文将基于组的轨迹分析模型用于分析文献的引用轨迹,对引文轨迹进行拟合分组,使我们对于引文的研究不仅仅局限于累计被引频次,而在引用行为和模式方面有更深的了解。然而,这种方法也有一定的局限性。对于组数的判断和每个亚组的拟合方程,仍需要根据主观经验来判断并根据结果选取最优的模型。此外,当系统较为复杂、数据量较大、存在异常值或者模型较为复杂时,拟合可能存在不收敛的情况。因此,GBTM 的作用仅仅是提供了一种在一定条件下研究较长时间内引文轨迹的方法。

其次,本文虽然在一定范围内对引文轨迹进行了拟合,但是忽略了单篇论文的轨迹特点。在文献的被引生命周期中,可能会出现多个引文高峰。然而,由于拟合分组的过程主要提取共性特征,个体的特征被削弱,使其在本文的研究中不能够完全体现。但是,引文轨迹分组也提供了更加细粒度地研究引文模式的可能性,通过观察不同类型的轨迹,可以针对性地对某一组的文献特征进行更深入的挖掘。

此外,引文轨迹只能反映出论文的直接被引情况,即只考虑了文献的直接影响力,而文献产生的影响并非总是直接的。X. Hu 等[44]的研究指出,直接被引数量有时并不能反映出论文的真实影响力,如有些文献的影响力体现在其产生的"触发作用",引发了其他论文被引用。因此,要全面评估文献的影响力,仍有必要将间接引用和引文级联结构纳入考量范围。

参考文献:

[1] BORNMANN L, MUTZ R, NEUHAUS C, et al. Use of citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results[J]. Ethics in dcience and environmental politics, 2008, 8(1):93-102.

[2] LINE M B, SANDISON A. "Obsolescence" and changes in the use of literature with time[J]. Journal of documentation, 1974, 30(3):283-350.

[3] AMIN M, MABE M A. Impact factors: use and abuse[J]. Medicina (Buenos Aires), 2003, 63(4):347-354.

[4] AVRAMESCU A. Actuality and obsolescence of scientific literature [J]. Journal of the American Society for Information Science, 1979, 30 (5):296 - 303.
[5] 李江，姜明利，李玥婷. 引文曲线的分析框架研究——以诺贝尔奖得主的引文曲线为例[J]. 中国图书馆学报，2014，40(2):41 - 49.
[6] VAN RAAN A F J. Sleeping beauties in science[J]. Scientometrics, 2004, 59(3):467 - 472.
[7] LANGE L L. Sleeping beauties in psychology: comparisons of "hits" and "missed signals" in psychological journals[J]. History of psychology, 2005, 8(2):194 - 217.
[8] MINGERS J. Shooting stars and sleeping beauties: the secret life of citations[C]//The 22nd European conference on operational research(Euro XXII). Prague: Kent Academic Repository, 2007:8 - 11.
[9] VAN DALEN H P, HENKENS K. Signals in science-on the importance of signaling in gaining attention in science[J]. Scientometrics, 2005, 64 (2):209 - 233.
[10] 周晓英，董伟，朱小梅，等. 图书馆学情报学高影响力论文特征及所反映的学科差异分析[J]. 中国图书馆学报，2012，38(4):99 - 109.
[11] VANCLAY J K. Factors affecting citation rates in environmental science[J]. Journal of informetrics, 2013, 7 (2):265 - 271.
[12] 王剑，高峰，王健，等. 被引频次与引用认知相关性的实证研究[J]. 图书情报工作，2014，58 (13):95 - 99.
[13] 牛昱昕，宗乾进，袁勤俭. 开放存取论文下载与引用情况计量研究[J]. 中国图书馆学报，2012，38 (4):119 - 127.
[14] 成颖. 面向学术新人的相关性判据研究——基于本科课程论文的内容分析[J]. 情报学报，2011，30 (9):922 - 934.
[15] BORNMANN L, LEYDESDORFF L, WANG J. How to improve the prediction based on citation impact percentiles for years shortly after the publication date? [J]. Journal of informetrics, 2014, 8 (1):175 - 180.
[16] 闵超，DING Y，李江，等. 单篇论著的引文扩散[J]. 情报学报，2018，37 (4):341 - 350.
[17] FU L D, ALIFERIS C F. Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature[J]. Scientometrics, 2010, 85 (1):257 - 270.
[18] ROLDAN-VALADEZ E, RIOS C. Alternative bibliometrics from impact factor improved the esteem of a journal in a 2-year-ahead annual-citation calculation: multivariate analysis of gastroenterology and hepatology journals[J]. European journal of gastroenterology and hepatology, 2015, 27 (2):115 - 122.
[19] BORDONS M, MORILLO F, GOMEZ I. Analysis of cross-disciplinary research through bibliometric tools[M]//Handbook of quantitative science and technology research. Dordrecht: Springer Netherlands, 2005:437 - 456.
[20] LEVITT J M, THELWALL M. The most highly cited Library and Information Science articles: interdisciplinarity, first authors and citation patterns[J]. Scientometrics, 2009, 78(1): 45 - 67.
[21] KWON S, SOLOMON G E A, YOUTIE J, et al. A measure of knowledge flow between specific fields: implications of interdisciplinarity for impact and funding[J]. PLoS ONE, 2017, 12(10): e0185583.
[22] ABRAMO G, D'ANGELO C A, COSTA F D. Do interdisciplinary research teams deliver higher gains to science? [J]. Scientometrics, 2017, 111(1): 317 - 336.
[23] BHAT H S, HUANG L H, RODRIGUEZ S, et al. Citation Prediction Using Diverse Features[C]// 2015 IEEE International Conference on Data Mining Workshop. Atlantic: IEEE, 2015: 589 - 596.
[24] PONOMAREV I V, LAWTON B K, WILLIAMS D E, et al. Breakthrough paper indicator 2.0: can geographical diversity and interdisciplinarity improve the accuracy of outstanding papers prediction? [J]. Scientometrics, 2014, 100(3): 755 - 765.
[25] RINIA E J, VAN LEEUWEN T N, VAN RAAN A F J. Impact measures of interdisciplinary research in physics[J]. Scientometrics, 2002, 53(2): 241 - 248.
[26] LEVITT J M, THELWALL M. Is multidisciplinary research more highly cited? a macrolevel study[J]. Journal of the American Society for Information Science and Technology, 2008, 59(12): 1973 - 1984.
[27] WANG J, THIJS B, GLÄNZEL W. Interdisciplinarity and impact: distinct effects of variety, balance, and disparity[J]. PLoS ONE, 2015, 10(5): e0127298.
[28] 苏新宁. 提升图书情报学学科地位的思考——基于 CSSCI 的实证分析[J]. 中国图书馆学报，2010，36(4):47 - 53.
[29] MINGERS J. Exploring the dynamics of journal citations: modelling with S-curves[J]. Journal of the Operational Research Society, 2008, 59(8):1013 - 1025.
[30] MINGERS J, BURRELL Q L. Modeling citation behavior in management science journals[J]. Information processing & management, 2006, 42(6):1451 - 1464.
[31] PILKINGTON A. Modeling citation diffusion: innovation management literature[J]. International journal of innovation and technology management, 2013, 10(1):1350004.
[32] BOUABID H. Revisiting citation aging: a model for citation distribution and life-cycle prediction[J]. Scientometrics, 2011, 88 (1):199 - 211.
[33] NADARAJAH S, KOTZ S. Models for citation behavior[J]. Scientometrics, 2007, 72(2):291 - 305.
[34] NAGIN D S. Group-based modeling of development[M]. Cambridge, MA: Harvard University Press, 2005.
[35] NAGIN D S, ODGERS C L. Group-based trajectory modeling (nearly) two decades later[J]. Journal of quantitative criminology, 2010, 26(4), 445 - 453.
[36] BOUABID H. Revisiting citation aging: a model for citation distri-

bution and life-cycle prediction[J]. Scientometrics, 2011, 88(1):199–211.
[37] PORTER A L, ROESSNER J D, COHEN A S. Interdisciplinary research: meaning, metrics and nurture[J]. Research evaluation, 2006, 15(3):187–195.
[38] CARLEY S, PORTER A L. A forward diversity index[J]. Scientometrics, 2011, 90(2):407–427.
[39] STIRLING A. A general framework for analysing diversity in science, technology and society[J]. Journal of the Royal Society Interface, 2007, 4(15):707–719.
[40] AVERSA E S. Citation patterns of highly cited papers and their relationship to literature aging: a study of the working literature[J]. Scientometrics, 1985, 7(3):383–389.
[41] BAUMGARTNER S E, LEYDESDORFFe L. Group-based trajectory modeling of citations in scholarly literature: dynamic qualities of "Transient" and "Sticky" knowledge claims[J]. Journal of the Association for Information Science and Technology, 2013, 65(4):797–811.
[42] MCFASSEN D, REID F. Aggregate travel demand forecasting from disaggregated behavioral models[J]. Transportation research record: travel behavior and values, 1975, 534(534):24–37.
[43] MCFASSEN D. Regression-based specification tests for the multinomial logit model[J]. Journal of econometrics, 1987, 34(1):63–82.
[44] HU X, ROUSSEAU R. Scientific influence is not always visible: The phenomenon of under-cited influential publications[J]. Journal of informetrics, 2016, 10(4):1079–1091.

作者贡献说明:
石燕青:负责研究框架设计与论文撰写;
赵一方:负责数据处理和调试。

The Influence of Interdisciplinarity on Citation Trajectory in Scientific Literature

Shi Yanqing[1] Zhao Yifang[2]
[1] College of Information Science & Technology, Nanjing Agriculture University, Nanjing 210095
[2] Information Management School, Nanjing University, Nanjing 210023

Abstract: [**Purpose/significance**] This paper attempt to investigate differences among citation trajectories in subpopulations of articles in the field of library and information science, and analyze the underlying factors. [**Method/process**] Group-based trajectory modeling (GBTM) was applied to the citation curves of articles to distinguish among the development trajectories. The interdisciplinarity were measured to investigate the relationship between the interdisciplinary degree and the citation trajectory. [**Result/conclusion**] There are two citation modes, "persistent" and "temporary", at different cumulative cite levels. As a predictor variable, interdisciplinarity do not affect the levels of aggregated citations but only the persistence of citations.
Keywords: citation anaysis citation trajectory citation mode interdisciplinary research

勘误说明

《图书情报工作》2018 年第 62 卷第 20 期第 49–56 页的文章《国内高校图书馆同伴教育服务现状调查与发展策略研究》,部分内容表述有误,现更正如下:

原文第 54 页“高校图书馆的同伴教育往往衍生于其所在高校已经开展同伴教育项目,如首都师范大学图书馆的同伴教育项目就是在该校研究生院主办的同伴教育项目取得比较好的效果后才开展的,在开展同伴教育活动时为了扩大宣传渠道,借助于研究生院的宣传平台进行宣传。”更正为“高校图书馆的同伴教育往往在其所在高校已经开展新生教育、职业规划等同伴教育项目的情况下,利用同伴教育理念结合图书馆的工作进行适合图书馆的同伴教育服务,并和学校其他部门的同伴教育项目开展合作。”

特此更正!

《图书情报工作》编辑部